

Internal errors: a valid alternative for clustering estimates?

Pablo Arnalte-Mur and Peder Norberg

Institute for Computational Cosmology, Durham University
South Road, Durham DH1 3LE, United Kingdom
email: pablo.arnalte-mur@durham.ac.uk

Abstract. We investigate the validity of internal methods to estimate the uncertainty of the galaxy two-point correlation function. We consider the jackknife and bootstrap methods, which are based on re-sampling sub-regions of the original data. These are cheap computationally, and do not depend on the accuracy of external simulations. We test the different methods over a large range of scales using a set of 160 mock catalogues from the LasDamas set of simulations. Our results show that the standard bootstrap method significantly overestimates the true uncertainty at all scales. We try two possible generalisations of the bootstrap, but find them not to be robust. Regarding jackknife, we obtain that this method provides an unbiased estimation of the error at small and intermediate scales, up to $\sim 40 h^{-1}$ Mpc. At larger scales, it typically overestimates the error by a $\sim 13\%$.

Keywords. methods: data analysis, methods: statistical, galaxies: statistics, large-scale structure of universe

1. Introduction

The study of galaxy clustering in large redshift surveys through the measurement of the two-point correlation function (2PCF) provides valuable information about the global properties of the Universe and about the process of formation and evolution of galaxies. It is important to understand the different methods available to accurately estimate the uncertainty of these measurements.

One option is to use an external estimate: generate a large number of simulated realisations of the data, and compute the uncertainty from the distribution of the measurements in them (Dodelson & Schneider 2013; Taylor *et al.* 2013). An alternative is to use internal error estimations, based on the re-sampling of the data itself. Different methods exist, such as jackknife and bootstrap methods (Efron & Tibshirani 1993). These methods are very cheap computationally and, as they use the data directly, do not depend on the accuracy of external simulations. However, their bias and variance are not very well understood for the case of clustering statistics.

In this work, we test the validity of some of these internal methods for a typical galaxy sample. We focus on the errors of the 2PCF over a large range of scales.

2. Resampling methods

We focus on the study of the redshift-space 2PCF estimated using the Landy & Szalay (1993) estimator, which is based on counting pairs of galaxies in both the data sample and in an auxiliary Poisson catalogue containing the same selection effects:

$$\hat{\xi}(s) = 1 + \frac{DD(s)}{RR(s)} - 2 \frac{DR(s)}{RR(s)}, \quad (2.1)$$

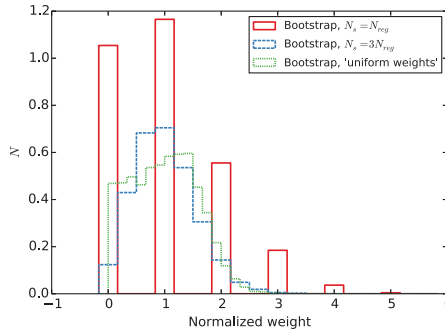


Figure 1. Distribution of weights w_i^k for different bootstrap methods, for the case $N_r = 10$. The weights are normalised such that $\sum_i^{N_r} w_i^k = 1$ for each realisation k .

where the different symbols correspond to the suitably-normalised pair counts in the data catalogue ($DD(s)$), in the random catalogue ($RR(s)$), and to the cross-pairs between both catalogues ($DR(s)$), as function of pair separation.

The internal error methods used in Cosmology (see Norberg *et al.* 2009 for a review) are based on dividing the data in a set of N_r non-overlapping regions or ‘blocks’. New ‘mock realisations’ are created by resampling these blocks, the 2PCF for each realisation, ξ^i , is computed using eq. (2.1) and the uncertainty is computed from the distribution of ξ^i . The resampling is based on blocks instead of individual data points to take into account the fact that the points themselves are correlated.

In the standard bootstrap method, each realisation is created by selecting randomly N_r blocks with replacement from the data. The uncertainty is estimated directly from the variance of the ξ^i . In the jackknife method, N_r realisations are created by omitting in each case one of the blocks from the data. The uncertainty is now obtained by multiplying the variance of the ξ^i by a factor of $(N_r - 1)$, to take into account the large overlap between realisations constructed in this way.

Generalisation of the bootstrap method. Norberg *et al.* (2009) noted that the realisations generated using the standard bootstrap method encompass only $\sim 60\%$ of the total volume covered by the data, inducing a systematic overestimation of the error. They proposed to generalise the bootstrap method by selecting in each realisation N_s blocks from the N_r available, with $N_s > N_r$. An equivalent approach is to regard the bootstrap procedure as a re-weighting of the ‘partial pair counts’ $DD_{ij}(s)$: the number of pairs of galaxies with separation s in which one of the galaxies is in region i and the other in region j . The total pair counts for a given bootstrap realisation k is then obtained as

$$DD^k(s) = \sum_i^{N_r} \sum_j^{N_r} w_i^k w_j^k DD_{ij}(s), \tag{2.2}$$

where w_i^k is the normalised weight assigned to region i in bootstrap realisation k . This same procedure is repeated for DR^k and RR^k and ξ^k is estimated using eq. (2.1).

Fig. 1 shows the distribution of weights that result from applying the bootstrap method with either $N_s = N_r$ (the standard case) or $N_s = 3N_r$. This approach allows us to investigate the possibility of generating bootstrap realisations using different weight distributions. As an example of this possibility, we test the case in which the weights for each block are drawn randomly from a uniform distribution. The resulting distribution of normalised weights w_i^k is also shown in Fig. 1.

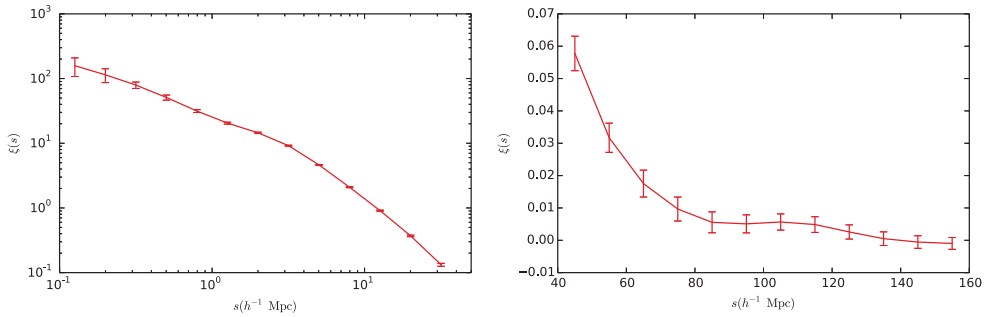


Figure 2. Mean 2PCF $\xi(s)$ measured from the 160 mock LasDamas catalogues at small (left) and large (right) scales. The error bars show the standard deviation of the measurements $\sigma_T(s)$, and correspond to the uncertainties for a single realisation.

3. Design of tests

We tested the different methods using a set of mock galaxy catalogues from the public LasDamas simulations[†] (McBride *et al.* in prep.). We used 160 independent realisations of their Luminous Red Galaxies (LRGs) sample, which reproduces the properties of the SDSS-DR7 catalogue ($V = 6.6 \times 10^8 \text{Mpc}^3 h^{-3}$, $N_{gal} \simeq 75000$).

We study the 2PCF for these catalogues using logarithmically-spaced bins in the range $0.1 < s < 40 h^{-1} \text{Mpc}$, and linearly-spaced bins in $40 < s < 160 h^{-1} \text{Mpc}$. From the different realisation's results, we compute the standard deviation of the $\xi(s)$ values in each bin, $\sigma_T(s)$, which we take as our 'true' uncertainty to use as reference for the internal error estimates. We plot the 2PCF of the mock catalogues in Fig. 2

For our resampling tests, we divide the survey in $N_r = 100$ regions, using a 10×10 grid in the RA, DEC coordinates. We have tested that our results do not change significantly with the number or geometry of the regions considered.

4. Results

We computed the 2PCF for each of the 160 mock realisations, and estimated its uncertainty in each case using the four methods described above: jackknife, standard bootstrap (with $N_s = N_r$), bootstrap with a larger number of resampled blocks ($N_s = 3N_r$) and bootstrap with uniformly distributed weights. Fig. 3 shows the estimation of the error σ_ξ we obtain in each case, as function of scale, compared with our reference value σ_T .

We find that the standard bootstrap method overestimates the error at all scales by $\sim 40 - 60\%$, in agreement with previous results (e.g. Norberg *et al.* 2009). We tested two generalisations of the bootstrap method that aimed at correcting this effect. At first sight, these work better than the standard bootstrap, reducing the bias in the estimation to $7 - 25\%$. However, they end up underestimating the errors in most cases. Moreover, these results are not robust to changes in the N_s used or the continuous distribution used for the weights.

Jackknife is clearly the method giving the best results. At small scales ($s < 40 h^{-1} \text{Mpc}$), the bias in the estimated error is very small ($\sim 0.5\%$). This can be explained by the fact that at these scales most of the pairs considered correspond to galaxies in the same region, so we are close to the conditions in which the standard jackknife was shown to be valid (Efron & Tibshirani 1993). We note, however, that even in this case there is an important dispersion ($\sim 10\%$) of the values obtained for different realisations.

[†] <http://lss.phy.vanderbilt.edu/lasdamas/>

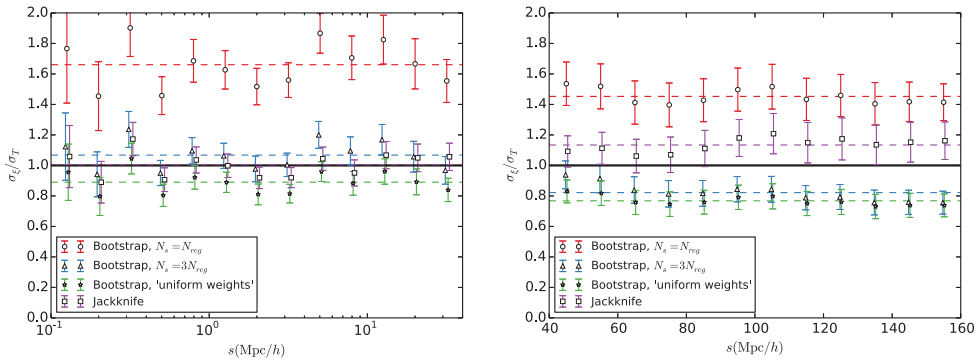


Figure 3. Errors estimations obtained using different internal methods applied to the LasDamas mocks, divided by the ‘true’ uncertainty (obtained from the variance of the realisations). The points and errorbars correspond to the mean and standard deviation of the errors estimated from the 160 realisations.

At larger scales, the estimator is biased, overestimating the error by a $\sim 13\%$, in agreement with the results of Guo *et al.* (2013). This could be caused by the significant fraction of ‘crossed pairs’ coming from galaxies in different regions at these scales. The fraction of ‘auto’ and ‘cross’ pairs as function of scale can be directly computed from the data, and thus could be used as a way of assessing the range of validity of the jackknife method.

5. Conclusions

We studied the performance of different internal methods to estimate the uncertainty on the galaxy 2PCF on a set of 160 realistic simulations. We also introduced a new approach to generalise the bootstrap method, based on the weighting of the different blocks.

Our results show that the standard bootstrap systematically overestimates the 2PCF uncertainty. The generalised method is not reliable, as it depends on an arbitrarily chosen distribution of weights. We found that the jackknife method provides an unbiased estimation of the error at small scales, while slightly overestimating the error at large scales. At these small scales, cosmological simulations could have more difficulties to reproduce the galaxy clustering pattern, so the external error estimation methods are less reliable. We consider therefore that the jackknife method can be very useful for clustering studies in this regime.

References

- Dodelson, S. & Schneider, M. D. 2013, *Phys. Rev. D*, 88, 063537
 Efron, B. & Tibshirani, R. J. 1993, *An Introduction to the Bootstrap*. Chapman & Hall/CRC, Boca Raton
 Guo, H., *et al.* 2013, *ApJ*, 767, 122
 Landy, S. D. & Szalay, A. S. 1993, *ApJ*, 412, 64
 Norberg, P., Baugh, C. M., Gaztañaga, E., & Croton, D. J. 2009, *MNRAS*, 396, 19
 Taylor, A., Joachimi, B., & Kitching, T. 2013, *MNRAS*, 432, 1928